Journal of Tekirdag Agricultural Faculty
Tekirdağ Ziraat Fakültesi Dergisi

Aralık/December 2023, 20(4)
Başvuru/Received: 22/08/22
Kabul/Accepted: 04/02/23
DOI: 10.33462/jotaf.1165105

http://dergipark.gov.tr/jotaf
http://jotaf.nku.edu.tr/

ARAŞTIRMA MAKALESİ **RESEARCH ARTICLE**

Image Processing and Traditional Machine Learning Based Classification of Brown Marmorated Stink Bug (*Halyomorpha Halys*) Defected Hazelnut*

Görüntü İşleme ve Geleneksel Makina Öğrenmeye Dayalı Fındıkta Kahverengi Kokarca (*Halyomorpha Halys*) Zararının Sınıflandırılması*

Omsalma Alsadig Adam GADALLA[1*], Yeşim Benal ÖZTEKİN[2]

Abstract

Quality control of hazelnuts is a major concern in many regions across the world, but particularly in Turkey as the world's largest hazelnut producer. Using image processing and deep learning techniques, this study intended to detect and classify healthy hazelnuts and hazelnuts infected with the Brown Marmorated Stink Bug. Infected hazelnut samples were collected from the 2021 production period by experts. A Guppy Pro CCD camera-based image acquisition system was used to capture hazelnut images. A total of 400 RGB hazelnut images were captured to train machine learning models. Image segmentation process was carried out to subtract hazelnut images from the background using the Thresholding technique. Moment features were extracted from RGB and l*a*b* spaces to be used to train traditional machine learning models. Furthermore, the most relevant and discriminative feature set was selected using the Boruta feature selection method. Traditional machine learning models including Random Forest, Support Vector Machine, Logistic Regression, Naive Bayes, and Decision Tree were trained twice, once with all features and another with the selected feature set only. The overall accuracy, statistical characteristics of the confusion matrix, and model training time were all calculated to evaluate and compare models performances. As a result, threshold value of 50 was determined from the gray level histogram and was able to separate hazelnut image from the background perfectly. Only seven moment features were identified as the most discriminative features out of 24 features. The SVM model with all feature vectors had the greatest classification accuracy of 98.75 %. When only the selected features were employed, the performance of Random Forest and Logistic Regression models improved to 97.5 and 96.25 %, respectively.

Keywords: Support vector machine, Hazelnut, Feature selection, Feature extraction, Boruta

[1]***Sorumlu Yazar/Corresponding Author:** Omsalma Alsadig Adam, Gadalla, Department of Agricultural Engineering, Faculty of Agriculture, Khartoum University, Khartoum, Sudan. E-mail: omsalmagadalla@gmail.com OrcID: 0000-0001-6132-4672
[2]Yeşim Benal, Öztekin, Department of Agricultural Machinery and Technologies Engineering, Faculty of Agriculture, Ondokuz Mayis University, Samsun, Turkey, Samsun, Turkey. E-mail: yurtlu@omu.edu.tr OrcID: 0000-0003-2387-2322
Atıf/Citation: Gadalla, O. A. A., Öztekin, Y. B. (2023). Image processing and traditional machine learning based classification of brown marmorated stink bug (Halyomorpha Halys) defected hazelnut. *Journal of Tekirdağ Agricultrual Faculty,* 20(4): 784-798.

*This study was summarized from the PhD thesis.

Öz

Fındığın kalite kontrolü, dünyanın birçok bölgesinde, özellikle de dünyanın en büyük fındık üreticisi olan Türkiye'de büyük bir problem kaynağıdır. Bu çalışma, görüntü işleme ve derin öğrenme tekniklerini kullanarak, Kahverengi Kokarca ile enfekte olmuş ve sağlıklı fındıkları birbirinden ayırarak belirlemek ve sınıflandırmak amaçlanmıştır. Kahverengi Kokarcalı fındık örnekleri, uzmanlar tarafından 2021 üretim döneminden elde edilmiştir. Fındık görüntülerini yakalamak için Guppy Pro CCD kamera tabanlı görüntü alma sistemi kullanılmıştır. Geleneksel makine öğrenme modellerini eğitmek için toplam olarak 400 RGB fındık görüntüsü alınmıştır. Fındık görüntülerinin arka plandan çıkarılması için görüntü bölüntüleme işlemi Eşikleme tekniği kullanılarak gerçekleştirilmiştir. Fındık moment özellikleri, geleneksel makine öğrenme modellerini eğitmek için kullanılmak üzere RGB ve l*a*b* renk çıkarılmıştır. Ayrıca, Boruta özellik seçim yöntemi kullanılarak en önemli ve en ayırt edici öznitelik seti seçilmiştir. Rastgele Orman, Destek Vektör Makinesi, Lojistik Regresyon, Naive Bayes ve Karar Ağacı dâhil olmak üzere geleneksel makine öğrenme modelleri, bir kez tüm özelliklerle ve bir kez daha yalnızca seçilmiş özelliklerle olmak üzere iki kez eğitilmiştir. Genel doğruluk, karışıklık matrisinin istatistiksel özellikleri ve model eğitim süresinin tümü, modelin sınıflandırma performansını değerlendirmek ve karşılaştırmak için hesaplanmıştır. Sonuç olarak, gri seviye histogramından 50 eşik değeri belirlenmiştir ve fındık görüntüsünü arka plandan mükemmel bir şekilde ayırabilmiştir. Çıkartılmış 24 özellik arasından en ayırt edici özellik olarak sadece yedi tane renk özelliği belirlenmiştir. Tüm çıkartılmış özellikler kullandıktan sonra Destek Vektör Makinesi modeli kullanılarak, %98.75 ile en yüksek sınıflandırma doğruluğu elde edilmiştir. Aynı zamanda tüm özelliklerden sadece seçilen özellikler kullanıldığında Rastgele Orman ve Lojistik Regresyon (sınılandırıcılarının) modellerinin performansı sırasıyla %97.5 ve %96.25'e kadar yükselmiştir.

Anahtar Kelimeler: Destek vektör makinesi, Fındık, Özellik seçme, Özellik çıkartma, Boruta

1. Introduction

Hazelnut (*Corylus avellana* L.) is a seasonal fruit related to the family of Betulaceae of Fagales ordo (Aydinoglu, 2010). Hazelnut grows in various places of the world such as USA, Italy, China, and Spain, while Turkey produces the majority of commercial hazelnuts (FAOSTAT, 2019). Hazelnuts are extensively utilized in the confectionery industry due to their flavor and taste. They have a great nutritional value because they contain a variety of constituents, primarily lipids, carbohydrates, proteins, sugar, and dietary fibers (Memoli et al., 2017). However, many factors influence hazelnut production and quality around the world. Temperature, humidity, mold, and pests all have a significant impact on hazelnut quality and yield.

One of the most serious defects facing the production of hazelnuts in Turkey is hazelnut defection generated by the Brown Marmorated Stink Bug (BMSB). During the development period, the BMSB penetrates the hazelnut and feeds on the nuts. The damage that occurs inside hazelnut rises as the pest grows, and even a minor attack might jeopardize hazelnut quality. When the insect feeds, saliva is sent down one maxillary stylet while food is sucked up via another (Mitchell, 2018). Through the feeding process, the fruit's skin is damaged, and enclosing tissue is removed, resulting in serious injuries (Short, 2010). Depending on where the insect feeds on the fruits, they can produce blanks, shrink or corking incidents, or injuries on the kernels (Saruhan, 2010). Hazelnut kernel infected with the Brown Marmorated Stink Bug varies in their infection severity level. Due to the pest's different behaviors, the external symptoms could appear as very clear dark skin injuries and discoloration or very small entry points resulting from the insertion of the mouthpart (Short, 2010). The insect has not been managed yet, and the consequent loss from this insect is sometimes hidden or not visually detectable on the fruit surface (Saruhan, 2010). However, post-harvest hazelnut evaluations revealed that BMSB feeding appears to have harmed a significant proportion of nuts. The damaged kernels seem shriveled and withered, with evident feeding points consisting of a noticeable depression encircled by necrotic tissue areas that can lead to mildew and rotting (Molnar, 2010). Christopher (2014), evaluated the destruction caused by H. halys adults feeding on hazelnut kernels. As a result, the author claims that when stink bugs feed on ripe nuts, a higher percentage of kernels develop corky, white, rotting kernel tissue, with no statistically significant differences variation in the proportion of damaged kernels exhibiting these symptoms in field or laboratory studies. Economically, in the last few years; the pest caused damages of 200 Million Dollars in 2017, 300 Million in 2018, and about 500 Million in 2019 for the total hazelnut production in Turkey which negatively affected the hazelnut export value (Anonymous, 2019). Applying machine vision with pattern recognition techniques, machine learning, and deep learning algorithms to identify hazelnuts has numerous advantages over traditional sorting methods.

During post-harvest processing operations the primary method for detecting defects in the fruits is visual inspection, which includes sample preparation, consecutive sampling methods, visual and sensory characteristics examination, and dismisses nut classification. This technique, which is focused on visual and organoleptic evaluation takes time and trained workers for accurate nuts classification. Other methods have been studied to develop smart inspection techniques for inspecting hazelnut quality characteristics. Machines supported by mechanical approaches such as calibrator devices are used to characterize the dimensions of the shell and empty nuts. These machines are advantageous for the pre-treatment of the products, but they are not quality management mechanisms. A unique approach for detecting blank hazelnuts has been proposed, depending on an examination of the acoustic signal generated by the nut's action on a steel surface (Onaran et al., 2005).

In field of hazelnut defect detection, few works have been done using artifiecial intellegence based technologies such as; Deep learning , Machine learning, and Image processing. Fungal contaminated hazelnut kernels were categorized using A two-dimensional local discriminant bases algorithm and multispectral imaging technique. For hazelnuts that were both contaminated and uncontaminated with aflatoxin, a classification accuracy of 92.3% was attained (Kalkan and Çetisli, 2011). Solak and Altinisik (2018) classified and detected hazelnuts using image processing and clustering techniques. The size and area features were extracted from hazelnut images and hazelnuts were divided into three classes. The use of mean-based classification and K-means clustering algorithms yielded 100 detection and classification accuracy. The whole defective hazelnuts were detected by Kivrak and Gürbüz (2019) using image processing and machine learning techniques. The goal of the work was to distinguish intact hazelnuts from damaged or defective ones. Images of hazelnut samples were captured using a cell phone and processed using image labelling techniques. Satisfactory results were obtained using the supervised learning method. Furthermore, a computer vision system was used to classify the partly skin removed hazelnut kernel, skin removed and rotten hazelnuts kernels. The processed hazelnut kernels are classified with a classification accuracy rate of 93.57 % (Guvenc et al., 2015)

There is no satisfactory technique for detecting insect pests in hazelnuts rapidly after harvest. Under existing production procedures, the most common technique for detecting defected hazelnuts involves manual selection, which is time-consuming, subjective, labor-intensive, and does not recognize hazelnut with very low severity levels. As a result, there is a need to raise the level of subjectivity, stability and effectiveness in assessing hazelnut quality. The digital image processing Approaches and artificial intelligence methods can play a significant role in this endeavor as the method offers the potential for high speed, non-destructive classification of hazelnut (Yadhunath et al., 2022). Machine vision systems are suitable for inspecting rigid and predefined objects. However, visual characteristic of agricultural products such as color, shape, texture are difficult for machine vision system to discern. Artificial intelligence algorithms such as artificial neural network have presented to be powerful in dealing with the type of problems that require interpolation of huge amount of data. The main aim of this study was to apply image analysis and different machine learning algoriths to separate healthy hazelnut from BMSB-defected hazelnut.

2. Materials and Methods

2.1. Materials

2.1.1. Hazelnut Samples

For classification purpose, normal and Brown Marmorated Stink Bug-infected hazelnuts were used in this experiment. Brown Marmorated Stink Bug-infected hazelnut samples were taken from a commercial hazelnut processing factory, which were previously harvested in the Black Sea region of Samsun. Before using hazelnut samples in this experiment, specialist assessors from the Plant Protection Department identified the damaged hazelnut samples as Brown Marmorated Stink Bug-infested samples. Infected hazelnut samples contained a mixture of hazelnut varieties, however the most commonly observed type was the Tombul variety, which is the most widespread hazelnut variety in the Black Sea region. *Figure 1* shows hazelnut samples infested with brown marmorated stink bug at various levels of severity and healthy hazelnut samples.

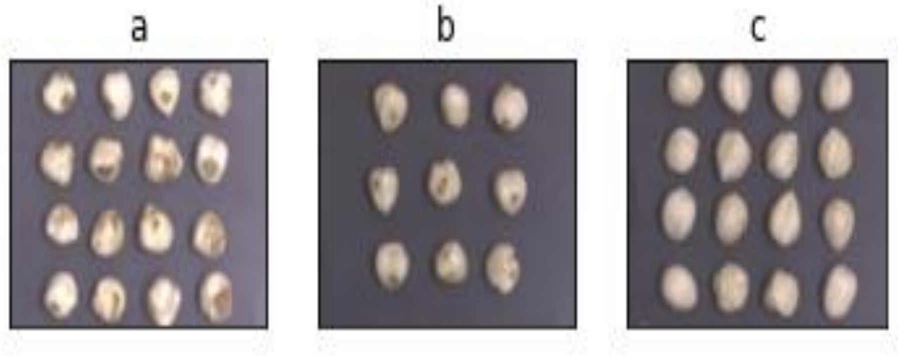

Figure 1. Normal and BMSB infected hazelnut samples, a; severely infected samples, b; slightly infected samples, c; normal samples

2.1.2. Computer Vision and Image Acquisition System

The image acquisition and machine vision system in this work was constructed and built in the biological material laboratory of Ondokuz Mayis University's Agricultural Machinery and Technologies Engineering Department (*Figure 2*). The components of the machine vision and image acquisition system were; darkened imaging chamber for adequate and well-distributed light, an image acquisition camera, a halogen lamp, a computer, and software.

Figure 2. Image acquisition system

2.1.2.1. Image Acquisition

For image acquisition, an Allied Vision Technology Guppy PRO F-032 economical FireWire camera with a Sony ICX424 CCD type sensor was employed. As demonstrated in *Figure 3* the sensor used for image acquisition was capable of capturing images in color and monochrome formats at wavelengths extending from 400 to 1000 nm at 82.0 frames p er second and 0.3 MP resolutions. The system showed high performance in RGB color space when used to classify cashew kernels using color features (Baitu et al., 2023).

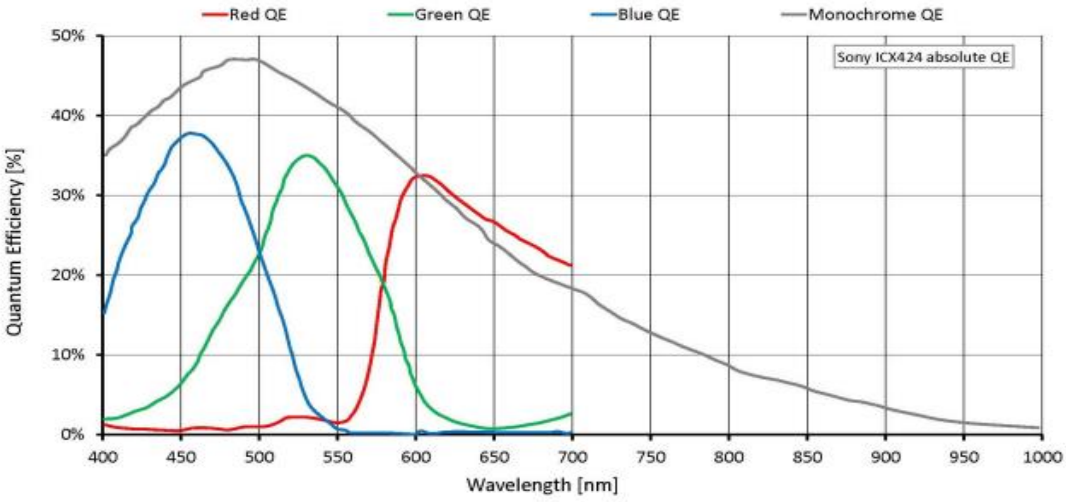

Figure 3. Absolute quantum efficiency of the CCD sensor in RGB and Monochrome color spaces

2.1.2.2. Image Processing and Image Analysis Software

In this study, Spyder (Python 3.8.8) was used as image analysis and image processing software (*Figure 4*), which is initially created and developed by Raybaut (2009) as an integrated development environment (IDE) for Python-based programming languages. Spyder software incorporates many well-known scientific Python packages. The packages that were used in this research included; NumPy for reading images in form of arrays, SciPy for solving mathematical, scientific, engineering, and technical problems besides manipulating and visualizing data, Matplotlib for plotting images and figures, Pandas for adjusting data in form of data frames, and other open-source software. On the other side, Open CV (Open Source Computer Vision Library) was used mainly as an open-source library for machine and deep learning models, and image processing purposes in this study.

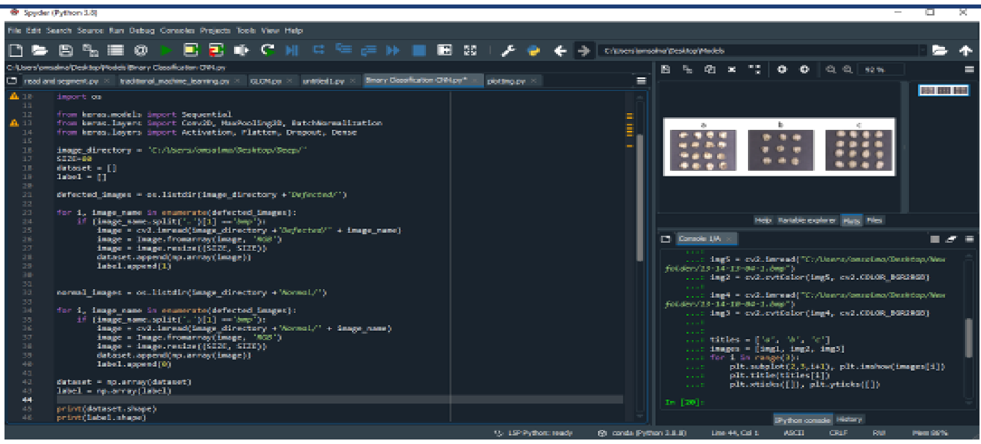

Figure 4. Spyder (Python 3.8.8) integrated development environment (IDE) Software

2. 2. Methods

2.2.1. Hazelnut Image Acquisition

Using the image acquisition system and the AVT Smart View image capture application, hazelnut samples were manually positioned on the sample holder and placed under the camera one by one. Two separate portions of the hazelnut surface were rotated and randomly captured for each sample. A total of 400 defective and normal hazelnut samples were captured in RGB (400-700) color space from a distance of 37 cm to be used in traing machine learning models as training and validation datasets. As the aim of the study was a classification target task, hazelnut images were captured at a resolution of 80x80 pixels.

2.2.2. Image Processing

2.2.2.1. Image Segmentation for Background Removal

Various image segmentation techniques are available, but the most applicable methods are; Otsu's thresholding, adaptive thresholding, and simple (global) thresholding techniques. Otsu thresholding technology is a powerful background extraction approach, but sometimes the algorithm uses large T values. Adaptive thresholding is a technique that calculates the threshold value for smaller regions, resulting in varying threshold values for different regions, making this technique unsuitable for our scenario. Global thresholding determines the threshold value based on the histogram of the overall pixel intensity distribution of the image as it performs better under controlled illumination conditions. Since in our case, it was capable to observe the image histogram clearly in the image acquisition software, global thresholding was the suitable method for separating hazelnut samples from the background. An image histogram is a graph that shows the distribution of image intensities in a given color space. In this study, a threshold value from the grayscale image histogram is used to transform hazelnut images from grayscale to binary images (black and white). The binary image was masked with the original image using the Bitwise-and operation to get the segmented RGB hazelnut images. The overall segmentation process is illustrated in *Figure 5*.

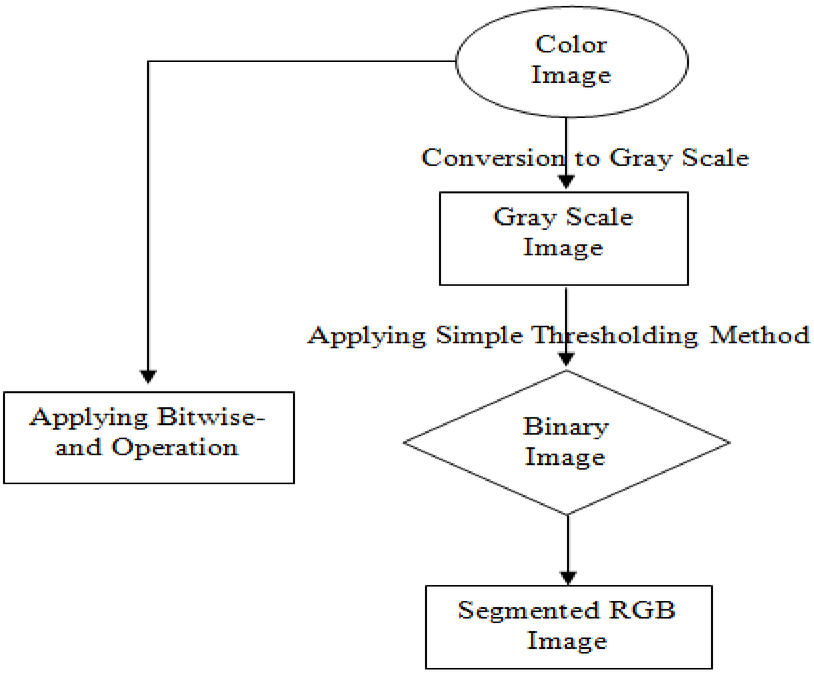

Figure 5. The overall process of hazelnut image segmentation algorithm

2.2.2.2. Feature Extraction

Proper feature extraction from the image is a key factor for a proper classification process. An appropriate feature set for a class should form a tight cluster around the centroid of that class. In this study, the insect performs shrivel or corking damage on hazelnut kernels and hazelnut kernel infected with the Brown Marmorated Stink Bug varies in their infection symptoms due to the pest's different behaviors. The external symptoms could appear as very clear dark skin injuries and discoloration or very small entry points resulting from the insertion of the mouthpart (Short, 2010). Also, changes in whole kernel color and changes in color where the mouth is inserted were observed (Ali, 2018). To improve the efficiency of hazelnut classification, color features from RGB and CIELAB (a color space specified by the International Commission on Illumination CIE) color spaces were extracted in this work. Therefore, hazelnut images in RGB and CIELAB (L*a*b*) color spaces were split into R (red), G (green), B (blue), L (l), A (a), and B (b). Moments features representing mean, variance, range, and skewness were captured from the split color spaces (Teimouri et al., 2016; Özlüoymak and Guzel, 2020).

2.2.2.3. Feature Selection

A feature selection algorithm was developed to select only the most relevant and discriminative features. Using BorutaPy from Boruta, a feature selector was created using the XGBoost classifier and utilized to fit in the train and test dataset. The contribution of all features was ranked and the features that ranked 1 were kept.

2.2.3. Machine learning Models

To determine the capability of RGB and lab color features in classifying BMSB-defected hazelnuts and healthy ones, the extracted features were used to train traditional machine learning classifiers. A total of 400 normal and defected hazelnut samples were used in this study for this purpose. 80% of the samples which count 320 hazelnut samples were applied for training purpose while the remaining 20% (80 samples) was used to test the classifiers. Here, the traditional machine learning classifiers were; Support Vector Machine, Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest classifier with 100 estimators. To evaluate the efficiency of the selected features in hazelnut classification, the classifiers were trained two times, firstly with the bulk features before applying any feature selection algorithm; secondly, the classifiers were trained using only the selected features. The model's performance was measured by calculating classification accuracy, confusion matrix and statistical parameters, and model execution (training) time.

2.2.4. Evaluation of Models Performances

For model performance evaluation, confusion matrix related to all traditional machine learning classifiers was extracted. The parameters in the confusion matrix express the actual labels of the hazelnut class and the labels predicted by the classifiers. The instances in the actual class are represented by the columns of the matrix, while the instances in the predicted class are represented by the rows. This can be used to see if the classifier is frequently mislabeling one item as another. A two-class confusion matrix is a two-row, two-column table that enables more detailed analysis than accuracy rate. The confusion matrix and the data entries for two-class classifiers are presented in *Table 1*. We used a confusion matrix in this work because classification accuracy is not a trustable measure for evaluating a classifier's efficiency, and it can generate misleading results when the quantity of samples in different classes varies greatly (Ghosh et al., 2014).

Table 1. Confusion matrix for two class classifier

	Predicted class	
	Positive	**Negative**
Actual class		
Positive	tp	fp
Negative	fn	tn

Where;

'tp' represents the number of positive samples classified as positive, 'fp' is the number of negative samples classified as positive, 'fn' is the number of positive samples classified as negative, and 'tn' is the number of negative samples classified as negative. Also, we evaluated the performance of our traditional models depending on some confusion matrix metrics such as precision, specificity, and sensitivity. The formulas of these measures are presented in *Table 2* (Taheri et al., 2015).

Table 2. Statistics parameters of confusion matrix and related formulas

Measure	**Formula**	**Evaluation focus**
Accuracy	$\frac{nTP\ nTN}{nTP\ nTN\ nFP\ nFN}$	Calculate the classifier's general performance.
Recall	$\frac{nTP}{nTP + nFN}$	The class arrangement between the data labels and the classifier's positive labels
Precision	$\frac{nTP}{nTP + nFP}$	An evaluation of a classifier's capacity to identify incidences of a specific class.
Specificity	$\frac{nTN}{nTN + nFP}$	How well a classifier can detect negative labels

3. RESULTS AND DISCUSSION

3.1. Image Segmentation

Image segmentation process is the primary stage of the image processing task. To segment hazelnut images from the background, simple (global) thresholding approach was applied. A threshold value of 50 was determined from the gray-level histogram (*Figure 6*) and the value was able to separate the background from hazelnut images properly. According to the histogram, hazelnut image pixels with intensities greater than 50 demonstrate the hazelnut image, while the pixels with intensities less than the threshold value demonstrate the dark background.

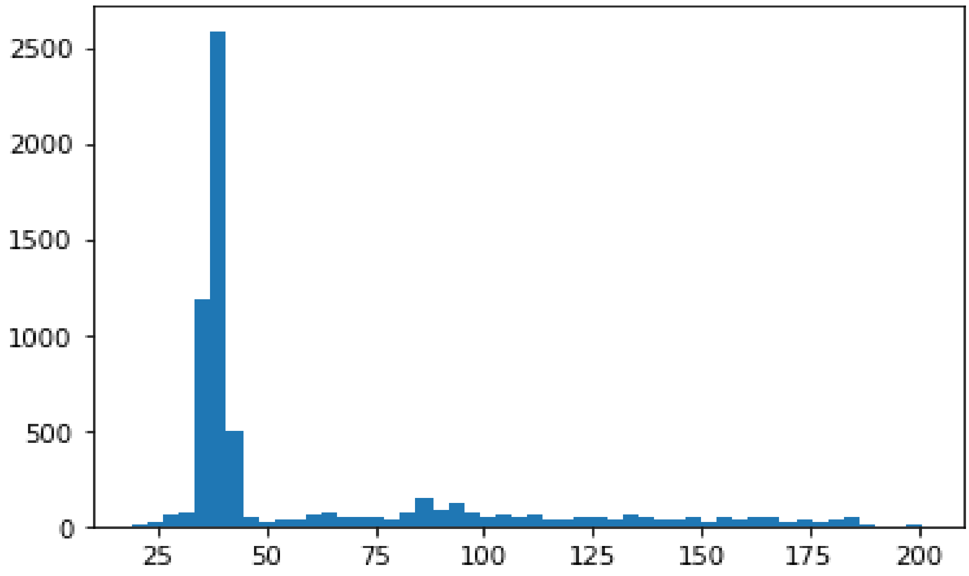

Figure 6. Gray-level histogram of hazelnut image

The applied background removal method was sufficient to get the required RGB hazelnut images as illustrated in *Figure 7*. The simple thresholding technique was efficient in the segmentation process and was applied frequently by researchers for background removal purposes (Senthilkumaran and Vaithegi, 2016).

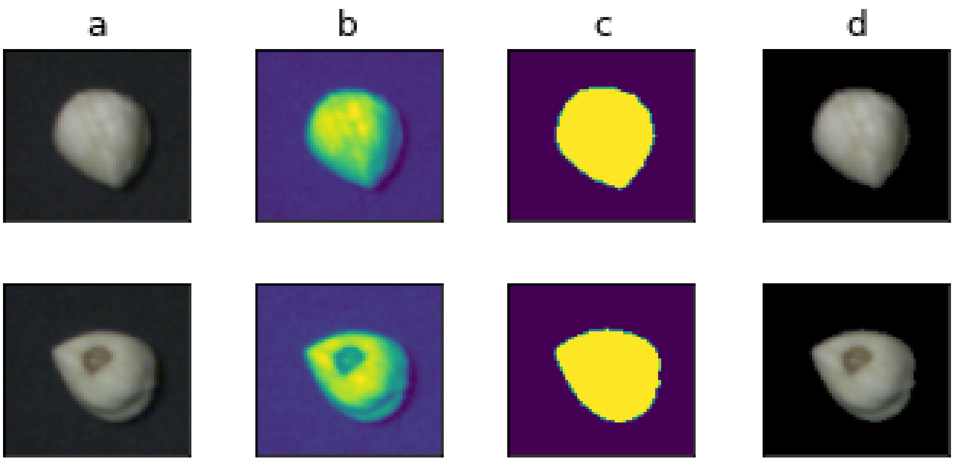

Figure 7. Segmentation steps. (a) RGB image, (b) Gray hazelnut image, (c) Thresholded image, (d) Segmented image

3.2. Feature Extraction and Selection

A total of 24 color features were extracted and saved to an excel file. These features were tested by the classifiers to evaluate their performance in predicting hazelnut classes. The applied feature selection algorithm confirmed that out of 24 features only 7 color features were important and discriminative more than the other features. The confirmed and rejected features are present in *Table 3*.

The majority of the accepted features were displayed in the l*a*b* color space, which reflects the reality described by Saruhan (2010) when researching the symptoms of BMSB infestation on hazelnut kernels after harvesting. In the l*, a*, and b* color spaces, the defection caused by the BMSB appears as discoloration and color changes. The characteristics of red mean, green skewness, and green mean were also found to be effective in classifying hazelnut into BMSB-defected and healthy hazelnut. This may support the theory of, BMSB infestations on hazelnut surfaces appear such little yellowish-green spots (Short, 2010).

Table 3. Confirmed and rejected features by applying BorutaPy

No.	Feature	Status	No.	Feature	Status
1	Red mean	Confirmed	13	l* mean	Rejected
2	Green mean	Confirmed	14	a* mean	Confirmed
3	Blue mean	Rejected	15	b* mean	Rejected
4	Red variance	Rejected	16	l* variance	Confirmed
5	Green variance	Rejected	17	a* variance	Rejected
6	Blue variance	Rejected	18	b* variance	Rejected
7	Red range	Rejected	19	l* range	Rejected
8	Green Range	Rejected	20	a* range	Rejected
9	Blue Range	Rejected	21	b* range	Rejected
10	Red skewness	Rejected	22	l* skewness	Rejected
11	Green skewness	Confirmed	23	a* skewness	Confirmed
12	Blue skewness	Rejected	24	b* skewness	Confirmed

3.3. Machine Learning Models Performance

Using test data, the performance of all trained traditional machine learning classifiers was assessed, as well as the confusion matrix. *Table 4* shows the classifier's performance resulting from training traditional machine learning classifiers with bulk features before any feature selection process.

Table 4. Classification accuracy and training time with bulk features

No.	Classifier	Acc. with bulk features	Training time
1	Support Vector Machine	98.75	0:00:00.007994
2	Logistic Regression	96.25	0:00:00.021987
3	Random Forest	95%	0:00:05.538398
4	Decision Tree	93.75	0:00:16.806870
5	Naïve Bayes	90%	0:00:00.001998

Based on all the extracted feature vectors, the highest classification accuracy (98.75%) was achieved from the Support Vector Machine model. This model can construct a good hazelnut classification system in real-time applications in food industries. Using the same color features, a Support Vector Machine-based computer vision system was developed and applied in real-time to classify pistachio nut kernels, and the same result was obtained (Nouri et al., 2017). Pacheco and López (2019) utilized RGB and l*a*b* color features with the digital image processing techniques for tomato flaws detection purposes. Multiple unsupervised machine learning algorithms were applied to identify the highest classification performance depending on the overall accuracy and confusion matrix metric and performance indices parameters such as precision, sensitivity, and specificity. Classification performances of 98.6, 98.3, 92, and 100% were achieved in the results for accuracy, specificity, precision, and sensitivity, respectively. These results consider very comparable to our results when RGB and l*a*b* color features were used. Acceptable performance was also achieved by training Random Forest and Logistic Regression models with bulk features, with an overall accuracy of 95 % and 96.25%, respectively. The confusion matrix was created to evaluate the effectiveness of each classifier in predicting individual hazelnut classes using all feature vectors as shown in *Figure 8*.

The Decision Tree and Naïve Bayes models had the lowest classification accuracy amongst the classifiers, with 90 % and 93.75 %, respectively. A comparable result was achieved from a Decision Tree-based classifier applied to classify areca nuts into healthy and defective nuts (Akshay and Hegde, 2021). Statistical parameters related to the confusion matrix were also calculated and presented in *Table 5*.

The Support Vector Machine showed the highest accuracy in predicting defective and healthy hazelnut samples when traditional machine learning classifiers were validated. However, the sensitivity of all classifiers (in except of Naïve Bayes classifier) in predicting defected hazelnut samples was very high and reached 100%. When predicting defected hazelnut samples. Individually, the least classification accuracy of defected hazelnut prediction was achieved from Naïve Bayes classifier. Hence, out of all defected hazelnut samples, 5 samples were falsely predicted as healthy samples which reduced the sensitivity of the model.

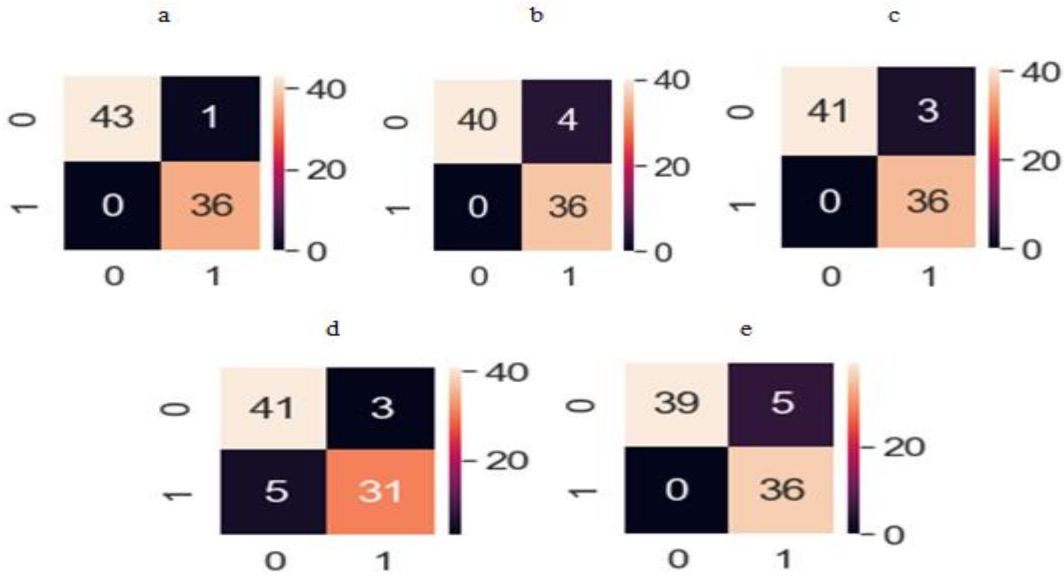

Figure 8. Confusion matrix of traditional classifiers; Support Vector Machine (a), Random Forest (b), Logistic Regression (c), Naïve Bayes (d), and Decision tree (e)

Table 5. Confusion matrix statistical parameters related to defected and healthy healthy classes with bulk features.

NO.	Measure %	Classifiers				
		SVM	RF	LR	NB	DT
1	Recall	100	100	100	89	100
2	Precision	97.7	90.9	93	93	88.6
3	Specificity	97	90	92	91	87.8

The classification accuracy of Random Forest and Naïve Bayes classifiers was improved by using the feature selection technique. That seems to be, the applied feature selection technique reduced the redundant data and minimized the opportunity of making decisions based on noise, resulting in increased accuracy (Balachandran et al., 2018). Furthermore, the Support Vector Machine's classification accuracy with the selected features was reduced to 96.25 %. *Table 6* shows the performance of traditional machine learning classifiers using the selected features.

Table 6. Classification accuracy and training time with selected feature

No.	Classifier	Acc. with bulk features	Training time
1	Random Forest	97.5%	0:00:00.256844
2	Support Vector Machine	96.25%	0:00:00.005997
3	Logistic Regression	96.25%	0:00:00.021984
4	Naïve Bayes	93.75%	0:00:00.001997
5	Decision Tree	93.75%	0:00:00.004996

Since the primary aim of the SVM classifier is to find the best and most appropriate separating hyperplane in an N-dimensional space that perfectly distinguishes the data point, in our case, using all input data (features) may have helped to create a plane with a maximum margin, which maximized the distance between the data points of the defected and healthy hazelnut classes, resulting in improved classification accuracy. When only selected features were used, all machine learning classifiers performed positively by decreasing model training time. *Figure 9* represents the confusion matrices of the classifiers with the selected features. Color features have been identified as powerful features in multiple classification tasks. Hong et al. (2011) used image processing and machine learning approaches to assess the quality of peanut products using the same color features. Based on pattern recognition, the color features R, G, and B of the damaged area were extracted to identify the damaged peanuts. The recognition task was completed with an accuracy of 80.12%. Furthermore, color features were selected from shape and color combinations to classify white cashew nut kernels using Artificial Neural Networks. from the input cashew kernel images, 24 color features were selected to achieve an accuracy of 88.93% (Narendra and Hareesha, 2016).

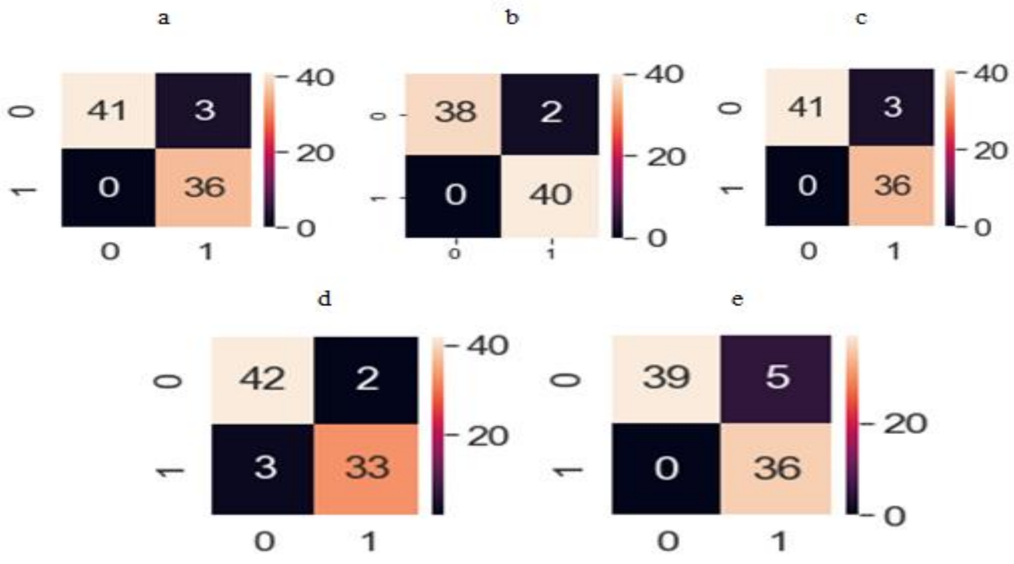

Figure 9. Confusion matrix of traditional classifiers; Support Vector Machine (a), Random Forest (b), Logistic Regression (c), Naïve Bayes (d), and Decision tree (e)

While the accuracy Random Forest and Naïve Bayes models was enhanced by considering only the relevant features, precision rate was also enhanced to be 95% for each of them. However, the statistical parameters of the SVM model were generally decreased as the overall classification accuracy was decreased as in *Table 7*.

Table 7. Confusion matrix statistical parameters related to defected and healthy classes with selected features.

NO.	Measure %	Classifiers				
		SVM	RF	LR	NB	DT
1	Recall	100	100	100	95	100
2	Precision	93	95	93	95	88.6
3	Specificity	92	95	92	94	87.8

4. Conclusion

After testing different background subtraction methods, the applied background subtraction thresholding method was able to extract the hazelnut image from the background perfectly and overcome the difficulties resulting from hazelnut spot color variation. The extracted color moment features represent in mean, variance, range, and skewness in the l*a*b* and RGB split color spaces were able to be describable and distinguishable in classifying healthy and hazelnut that defected with the BMSB. When all extracted moment features were applied for hazelnut classification, the Support Vector Machine classifier performed better than other traditional machine learning algorithms. After testing several feature selection approaches, the BorutaPy feature selector was identified as the most suitable one, and able to select the most relevant and discriminative features. Only seven features were selected from a total of 24 l*a*b* and RGB split color feature spaces. Using this strategy, data redundancy was also reduced. The majority of the features selected originated from the l*a*b* color spaces. As a result, the changes in hazelnut generated by brown marmorated stink insect infestation have been recognized to appear mostly in this color space after harvesting.

By training traditional machine learning classifiers with the selected futures only, the performances of most of these classifiers were improved. The performance of Random Forest, Naïve Byes, and Logistic Regression models was increased as well as the training time was reduced. In this case, the features selection objective was achieved by improving the learning process of machine learning models and increasing the predictive power of machine learning algorithms by selecting the most important and the most relevant variables and eliminating redundant and irrelevant features. There was no improvement in Decision Tress classifier performance in terms of accuracy, precision, and recall parameters. However, an improvement in processing time was achieved. The performance of the Support Vector Machine model by using only the selected features was decreased due to the margin maximization.

Although machine learning and deep learning models for computer and machine vision are widely used around the world and applied in various field, hazelnut research integrating this technology is very limited, leaving a significant gap in the sector. This gap is appearing in hazelnut processing factories and food industries. In addition, there are still manual and semi-automated hazelnut sorting and classification systems in use in post-harvest activities. These systems consider very poor, time-consuming, and inefficient systems. To overcome this challenge, it can be concluded that using the same technology more investigations are necessary for the hazelnut quality control to facilitate hazelnut processing easier in the food industry and increasing hazelnut quality. Hazelnut infestation with BMSB is a relatively new problem and is under investigation. In particular, hazelnut infestation by the BMSB poses a serious threat to hazelnut quality on a wide scale and negatively impacts national and international marketability and acceptability.

Acknowledgment

This work supported supported by Ondokuz Mayis University, Research Project No. PYO. ZRT.1904.21.001, Turkey.

References

Akshay, S. and Hegde, A. (2021). Detection and classification of areca nut diseases. 2021 *Second International Conference on Electronics and Sustainable Communication Systems* (ICESC). 1-3 September, P.1-5. Kottayam, India.

Ali, I. (2018). Hazelnut cultivation in Turkey. *Academic Journal of Agriculture,* 6(1): 1-13.

Anonymous. (2019). Fındıkta 'kahverengi- yeşil kokarca' tehdidi. https://www.sozcu.com.tr/2019/ekonomi/findikta-kahverengi-yesil-kokarca-tehdidi-4007632/ (Accessed date: December 2021).

Aydinoglu, A. C. (2010). Examining environmental condition on the growth areas of Turkish hazelnut (*Corylus colurna* L.). *African Journal of Biotechnology*, 9(39): 6492-6502.

Baitu, G. P., Gadalla O. A. A. and Oztekin, Y. B. (2023). Traditional machine learning-based classification of cashew kernels using colour features. *Journal of Tekirdağ Agricultural Faculty*, 20(1): 115-124.

Balachandran, P. V., Xue, D., Theiler, J., Hogden, J., Gubernatis, J. E. and Lookman, T. (2018). Importance of feature selection in machine learning and adaptive design for materials. *Materials Discovery and Design*, 4(4): 100-107.

Christopher. (2014). *The effects of kernel feeding by brown marmorated stink bug (Halyomorpha halys:Hemiptera: Pentatomidae) on commercial hazelnuts (Corylus avellana L.).* (MSc. Thesis). Oregon State University.

FAOSTAT. (2019). Crops Production Data. Retrieved from https://www. fao.org/faostat/en/data/QC (Accessed date: 10 April 2021).

Ghosh, S., Biswas, S., Sarkar, D. and Sarkar, P. P. (2014). A novel Neuro-fuzzy classification technique for data mining. *Egyptian Informatics Journal*, 15(3): 129-147.

Guvenc, S. A., Senel, F. A. and Cetisli, B. (2015). Classification of processed hazelnuts with computer vision. *23rd Signal Processing and Communications Applications Conference* (SIU)., 29-1 November, Germany.

Hong, C., Jing, W., Qiaoxia, Y. and Peng, W. (2011). Quality classification of peanuts based on image processing, *Huazhong Agricultural University Journal*, 24(5): 115-121.

Kalkan, H. and Çetisli, B. (2011). Online feature selection and classification. InAcoustics, Speech and Signal Processing (ICASSP), *2011 IEEE International Conference,* May 22 (pp. 2124-2127), Ankara, Turkey.

Kivrak, O. and Gürbüz, M. (2019). Detection of defective hazelnuts by image processing and machine learning. *Natural and Engineering Sciences,* 4(3): 100-106.

Memoli, A., Albanese, D., Esti, M., Lombardelli, C., Crescitelli, A., Di Matteo, M. and Benucci, I. (2017). Effect of bug damage and mold contamination on fatty acids and sterols of hazelnut oil. *European Food Research and Technology*, 243(4): 166-181.

Mitchell, P. L. (2018). Heteroptera as vectors of plant pathogens. *Neotropical Entomology*, 33:519–545.

Molnar, T. (2010). The brown marmorated stink bug: A new pest of hazelnut. *Journal of Food Technology*, 9(1): 33-38.

Narendra, V. G. and Hareesh, K. S. (2016). Prospects of computer vision automated grading and sorting systems in agricultural and food products for quality evaluation. *International Journal of Computer Applications,* 4(9): 75–87.

Nouri, A. H., Omid, M., Mohtasebi, S. S. and Firouz, M. S. (2017). Design, development and evaluation of an online grading system for peeled pistachios equipped with machine vision technology and support vector machine. *Information Processing in Agriculture*, 4(4): 333-341.

Onaran, I., Dulek, B., Pearson, T. C., Yardimci, Y. and Cetin, A. E. (2005). Detection of empty hazelnuts from fully developed nuts by impact acoustics. *2005 International Signal Processing Conference*, Sep 4 (pp. 1-4).

Özlüoymak, Ö, B. and Guzel, E. (2020). Determination of Colour and Kinetic Parameter Differences Between Aflatoxin Contaminated and Uncontaminated Pistachio Nuts Using Machine Vision. *Journal of Tekirdag Agricultural Faculty*, 18(1): 157-168.

Pacheco, W. D. N. and López, F. R. J. (2019). Tomato classification according to organoleptic maturity (coloration) using machine learning algorithms K-NN, MLP, and K-Means Clustering. *2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, Dec 30, (pp. 1-5) Bucaramanga, Colombia.

Raybaut, P. (2009). (Python) (ANN) Spyder v1.0.0 released". [Press release]

Saruhan, I. (2010). Brown Marmorated Stink Bug in Hazelnut (Palomena Prasina L. Heteroptera: Pe Tatomidae), *Anatolian Journal of Agricultural Sciences*, 25('2): 75-83.

Senthilkumaran, N. and Vaithegi, S. (2016). Image segmentation by using thresholding techniques for medical images. *Computer Science & Engineering: An International Journal*, 6(1): 1-13.

Short., B. A. (2010). Factors affecting appearance of stink bug (Hemiptera: Pentatomidae) injury on apple. *Environmental Entomology*, 39: 134-139.

Solak, S. and Altinisik, U. (2018). Detection and classification of hazelnut fruit by using image processing techniques and clustering methods. *Sakarya University Journal of Science*, 22(1): 56-65.

Taheri, G, A., Ahmadi, H., Omid, M., Mohtasebi, S. S., Mollazade, K., Smith, A. J. R. and Carlomagno, G. M. (2015). An intelligent approach for cooling radiator fault diagnosis based on infrared thermal image processing technique. *Applied Thermal Engineering Journal*, 87(4): 434-443.

Teimouri, N., Omid, M., Mollazade, K. and Rajabipour, A. (2016). An artificial neural network-based method to identify five classes of almond according to visual features. *Journal of Food Process Engineering*, 39(6): 625-635.

Yadhunath, R., Srikanth, S., Sudheer, A., Jyotsna, C. and Amudha, J. (2022). Detecting surface cracks on buildings using computer vision: an experimental comparison of digital image processing and deep learning. (197-210). In: Reddy, V. S., Prasad, V. K., Wang, J., Reddy, K. T. V. (eds) Soft Computing and Signal Processing. Advances in Intelligent Systems and Computing, vol 1340. Springer, Singapore.